DOCUMENT RESUME

ED 319 756 TM 014 855

AUTHOR Hummel, Thomas J.
TITLE Computer Assisted Instruction for Teaching Measurement-Based Clinical Hypothesis Testing.
PUB DATE Apr 90
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Clinical Diagnosis; *Computer Assisted Instruction; Computer Software Evaluation; *Counselors; Counselor Training; Cutting Scores; *Decision Making; *Hypothesis Testing; Postsecondary Education; Predictive Measurement; *Predictive Validity; Professional Education; Statistical Inference; Student Attitudes; *Test Use
IDENTIFIERS PASCAL Programing Language

ABSTRACT

A new computer-based approach for teaching counselors about predictive validity and how to use tests in clinical decision making was developed and tested. The approach is based on concepts similar to those used in inferential statistics, particularly statistical hypothesis testing. To implement this model, the counselor must: (1) determine the criterion variable and the rule for inclusion in the diagnostic category; (2) consider the consequences of misclassification; (3) set the maximum probability of misclassification allowable for any client; (4) compute a cutoff value on the predictor variable; (5) compare the client's score on the predictor variable to the cutoff value and make a decision; and, optionally, (6) determine long-run probabilities of true positives, false positives, true negatives, and false negatives that would result if the cutoff score was used over time with all clients. A computer program in PASCAL was written to teach this method. Students' opinions about the program were obtained from 26 counseling students. Students liked the program and found it helpful. Those who perceived themselves as relatively more able in statistics received the program more favorably than did those who saw themselves as less able. Five figures illustrate the study, and an appendix gives examples of student reactions. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

THOMAS J. HUMMEL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED319756

Computer Assisted Instruction for Teaching Measurement-Based Clinical Hypothesis Testing

By

Thomas J. Hummel
University of Minnesota

A Paper Presented at the Annual Meeting of the
American Educational Research Association
Boston, Massachusetts
April, 1990

TM014855

BEST COPY AVAILABLE

INTRODUCTION

Counseling psychologists in practice are familiar with the idea of assessing an individual client. As part of their daily work, they are less familiar with ways to assess the tests that might be a part of their client assessments. Predictive validity, i.e., the relationship between the test and later behaviors, thoughts, and feelings, is of paramount importance in assessing a test. For clinical use, predictive validity is arguably the most important attribute of a test. Test reliability is a necessary, but not sufficient, condition for a test's usefulness.

Counselor trainees' foundational work on predictive validity is often slanted toward using tests to meet institutional needs, such as college admissions or selecting candidates for jobs. Using tests for these "personnel decisions" is a well developed area. Concepts such as base rate and selection ratio, along with aids such as the Taylor-Russell tables (Taylor & Russell, 1939) help personnel psychologists think in a formalized way about selection decisions.

Another theme well represented in foundational work in testing is the use of tests to aid client decision making. In this context, a client is given expectancy data, such as the client's predicted GPA or probability of success at a given college. Here the client is the decision maker; the one who will act.

However, when counselors must diagnose or classify clients as a step in determining a course of action designed to meet a

client's needs, that is, when the counselor must act, the issues are somewhat different. The counselor's client might be at risk if the wrong course of action is taken. If the client's classification and associated treatment is to be determined using tests, then what decision model, if any, do counselors use in making classifications? It is the position of this paper that counselors are not well prepared in using formal decision models in diagnosing clients.

The purpose of the development and evaluation project described in this paper was to create and try out a new, computer-based approach for teaching counselors how to use tests in clinical decision making. The approach is based on concepts similar to those used in inferential statistics, particularly in statistical hypothesis testing.

To implement the model described in this paper for measurement-based hypothesis testing with clients, the counselor must carry out six essential steps. They are:

1. Determine the criterion variable and the rule for inclusion in the diagnostic category (e.g. what set of conditions must exist for a client to be considered "delinquent" or "depressed");
2. Consider the consequences of misclassification;
3. Set the maximum probability of misclassification allowable for any client (similar to setting an alpha level in statistical hypothesis testing);
4. Compute a cutoff value on the predictor variable (similar to a critical value in statistical hypothesis testing);
5. Compare the client's score on the predictor variable to the cutoff value and make a decision (similar to testing a statistical hypothesis by comparing a test statistic to a critical value); and optionally,
6. Determine the long run probabilities of true positives, false positives, true negatives, and false negatives which would result if the cutoff score were used over time with all clients (somewhat similar to Type I and Type II errors in statistical hypothesis testing).

If the counselor defines the criterion group, sets the maximum probability of misclassification, knows the relationship between predictor and criterion (e.g. validity coefficient), and is willing to assume a normal distribution for the continuous variables involved, then a cutoff on the predictor can be computed. Using the tables and formulas required is, at best, tedious, and it is unlikely that many counselor trainees would understand the concepts and rationale underlying the computations.

To make the above approach more understandable, a computer program was developed which integrates numerical analysis and computer graphics techniques to produce a sort of "visual calculator" for counselors to use in testing clinical hypotheses. Vivid color graphics and the ability to carry out "what-if" simulations by varying input to the program (e.g. validity coefficients) allow counselors to better understand the concepts, relationships, and issues which obtain when tests are used with clients.

While it is a relatively recent development that desktop computers have become capable of handling the computations involved, many of the ideas in clinical hypothesis testing have been in the literature for some time. The ideas expressed in Meehl and Rosen's (1955) seminal paper are foundational to the approach described here.

UNDERLYING ASSUMPTIONS

There are several assumptions that underlie the approach. They are:

1. Counselors can be viewed as professionals who diagnose and treat clients. This point of view is associated with the "medical model" and is rejected by some counselors. However, it is assumed here that counselors as human information processors

do develop categories of clients based on generalizations from experience, readings, and reflections. Further, it is assumed that, although counselors note exceptions to their general categories, categorizations do influence, if not determine, the behavior of the counselor. It is this process of categorization and modification of counselor behavior that is here viewed as diagnosis and treatment.

2. Misdiagnosis can have negative consequences. For example, viewing a client's lack of energy and negative mood states as symptomatic of depression, when in fact there is a physical cause (e.g., bacterial or hormonal), might cause a counselor to follow an inappropriate course of action.

3. Counselors should not harm their clients. Since the time of Hippocrates, helping professionals have felt it their duty to protect clients from harm. Counselors are better off working on the counselor-client relationship and developing a deeper understanding of the client than implementing a course of action for a wrong or nonexistent problem.

These three assumptions are important in determining which diagnostic error should be controlled, the "false positive" or the "false negative," and lead to a decision to control the probability of a false positive. Using an approach analogous to testing statistical hypotheses, one might test the "null" hypothesis that "the client is not depressed." A positive test would lead to the rejection of the hypothesis and to the conclusion that "the client is depressed." Having drawn that conclusion, treatment could begin for depression. A false positive in this situation is analogous to a Type I error in that the hypothesis would be rejected, but the person would not, in fact, be depressed. On the other hand, retaining as tenable the hypothesis of no depression when the person is, in fact, depressed, i.e., a false negative, is analogous to a Type II error. It is common practice in statistical hypothesis testing to control the probably of a Type I error. At times the

probability of a Type II error is considered and steps taken to reduce it. When the null hypothesis is retained in statistical hypothesis testing, one does not conclude that it is true, but merely that it has not been rejected and that one should not follow the course of action appropriate for the alternative hypothesis.

Analogously in clinical hypothesis testing, retaining the hypothesis that the client is not depressed would leave the client's depression an open question, and the counselor would continue attempting to classify the client. In this model, then, the counselor continues searching for a clear indicator of an appropriate treatment.

Once it is decided to control the probability of a false positive, the counselor must decide what level of error is acceptable given the consequences of an inappropriate treatment. It is assumed that for any given situation there is a maximum probability of error acceptable to the counselor. For example, the counselor may conclude that the probability of misclassifying any given client should be no higher than .25. In other words, the counselor would treat a client for depression only if people with the client's score on the predictor are depressed at least 75% of the time. Setting this probability is analogous to setting an "alpha level" in statistical hypothesis testing.

METHODS AND TECHNIQUES

This project had three major components: developing the computer program; integrating the program into the curriculum; and carrying out an initial evaluation. The program is written in Pascal and runs on IBM PC compatible computers with either EGA or VGA graphics. A fast microprocessor and a numeric coprocessor enhance the use of the program. While the numerical analysis and graphics techniques are standard, a special feature of the program is their integration. All graphs are scaled to represent accurately the results of computations carried out by the numerical analysis routines.

The program has two separate tracks, one for each of two models for the criterion variable. The first model conceptualizes the criterion variable as normally distributed, with the clinical group defined as those with criterion values above a defined cutoff. In this track, the user 1) names the criterion variable, 2) names the clinical group in the upper tail, 3) names those in the remainder of the distribution, 4) enters the mean and standard deviation of the criterion, 5) defines the cutoff value which separates the clinical group from the remainder of the distribution, 6) names the predictor, 7) enters the predictor's mean and standard deviation, 8) enters the correlation between criterion and predictor, and last, enters the maximum probability of misclassification allowed for any client. Given this input, the program computes the score on the predictor which has a conditional distribution yielding the required proportion of errors (Magnusson, 1967). Figure 1 shows the

graphics and text produced as a function of the preceding input. If one requests to see the long run consequences of using the computed predictor cutoff in classification, Figure 2 is produced, which displays the isodensity contours of a bivariate normal distribution and the regions associated with the various kinds of errors.

In the second program track, the criterion is conceptualized as a dichotomy, e.g., delinquent and not delinquent, with each category having normally distributed scores on the predictor. The user 1) enters the name of the larger of the two groups, 2) names the smaller, clinical, group, 3) names the predictor, 4) defines the mean and standard deviation of the larger group, 5) defines the mean and standard deviation of the clinical group, 6) enters how many times larger the larger group is than the smaller (base rate), and last 7) specifies the maximum probability of error. Figure 3 presents the resulting computer display and shows the predictor score for which the ratio of the likelihoods of the distributions results in the required proportion of errors. As in the first track, if one requests to see the long run consequences of using the computed cutoff in classification, Figures 4 and 5 are produced. They display the regions of the normal distributions associated with the various kinds of errors.

In both tracks, the user may change the input and determine what the consequences would be for the graphical and numerical results. For example, the impact of reduced predictive validity could easily be determined. This provides an opportunity for the

student to develop visual representations of the interrelationships among the various parameters of the models.

The program was designed for use in a course dealing with the clinical use of tests in counseling. Prerequisites for the course were introductory statistics and an introduction to tests and measurements. The teaching of clinical hypothesis testing took place in three phases, which comprised approximately one third of a three credit course. In the first phase, small groups of students discussed the consequences of misclassification, developed examples, and determined what level of classification error would be acceptable given the potential consequences. The small groups then shared their thinking in a class discussion and discovered the highest and lowest acceptable probabilities of error. Published validity studies were also introduced to give some idea of the range of validity coefficients which might be encountered in practice. The second phase took place in a computer classroom which contained 16 IBM PCs and a regular classroom equipped with a portable computer, liquid crystal display projection panel, and overhead projector. In this second phase, the students were introduced to clinical hypothesis testing and to the computer program. Both group and individual exercises were completed. In the third phase, held in a regular classroom, students worked on group activities dealing with responsibility in language. They began by reviewing phrases used in psychological reports (e.g. "seems to", "may tend to", "this suggests", and "might") and speculated about what level of confidence (or lack of confidence) the author was expressing. As

a last step, students discussed appropriate language to use with clients, given the fallibility of tests.

STUDENT REACTIONS

A student evaluation of the course was carried out using an anonymous student opinion survey. To this point, twenty-six students have used the program and seen multiple demonstrations of it.

Four items dealt specifically with student reactions to the program. All used a five point scale, with one equaling "strongly disagree" and five equaling "strongly agree." In responding to an item which dealt with how the program helped students "learn about the errors which occur when tests are used in decision making," the mean response was 4.27 and the standard deviation was .67. Over 96% of the students agreed or strongly agreed with the preceding statement.

An item dealing with the computer program's success in getting students to think had a mean of 4.39 and a standard deviation of .752. Over 92% of the students agreed or strongly agreed with this item.

The mean rating for an item concerning students' liking the program was 4.02. The standard deviation was 1.34. Approximately 73% of the students agreed that they liked the program, with the majority strongly agreeing.

The last of the four items dealing directly with the program was "Compared to other instructional aids used to teach statistical and measurement concepts, e.g., chalkboard, overhead

transparencies, the computer program was more effective." Slightly over 69% agreed or strongly agreed with this statement, while 11.5% disagreed or strongly disagreed. For this item, the mean was 3.81 and standard deviation was 1.06.

Three items dealt with characteristics of the students. Two of these items used the same five-point scale described above. The first was "I am very comfortable with computers." The mean response was 3.37 and the standard deviation was 1.21. The second was "I am very comfortable with statistical and measurement concepts." The mean and standard deviation for this item were, respectively, 2.96 and 1.30.

The third item was different with respect to its scale. This item asked students to compare themselves to other students with whom they had taken statistics and to estimate the percentage of those students that fell below them in achievement. The mean was 60.91 and the standard deviation was 18.67.

The items dealing with student characteristics were correlated with the ratings of the program. Only the last item, dealing with students' estimates of the percentage of students falling below them in statistics achievement, yielded statistically significant correlations. (The two-tailed, .05 level critical value for r was .433. This is the value for n = 21, the lowest n resulting when pairwise deletions were made to handle missing data.)

Significant correlations were found between the student achievement estimates and three ratings of the computer program. Achievement estimates correlated .504 with ratings of how the

program helped them learn about decision-making errors. A correlation of .471 was found between the estimates and students' liking of the program. Last, the correlation was found to be .566 between the achievement estimates and favorable comparisons of the computer program relative to other methods for teaching the material.

Students like the program and find it helpful in understanding the concepts being taught. There is some tendency for students who perceive themselves as relatively more able in statistics to view the computer program more favorably than those who see themselves as less able.

Anecdotally, many students were shocked to see the extreme cutoff scores necessary for maintaining acceptable levels of error with typical psychological instruments. With barely acceptable error levels, many were unnerved to see that a cutoff score was so extreme that only a handful of clients might be correctly diagnosed. (Examples of student reactions to the program are included in the Appendix.)

DISCUSSION

Teaching measurement-based clinical hypothesis testing with the aid of computers gives counselor trainees a solid grounding in concepts fundamental to the use of tests in client assessment. The approach makes clear the important limitations of currently available tests when they are used to select courses of action which are in the client's best interest. Further, the program gives vivid support to Goldman's (1972) concerns about tests and

counseling when he described their relationship as "The Marriage That Failed."

In addition to sensitizing trainees to test limitations, the approach presents a practical, yet formal, model for making clinical decisions. If this model were used with clients, counselors would be more likely to be tentative in their assessments, conscious of their responsibility in choosing language, and anxious to hear the clients reaction to hypothesis, suggested actions, and probabilities of success.

(To enhance the dissemination of this teaching approach, for at least one year following the presentation of this paper, faculty in higher education can receive a copy of the program at no charge by contacting the author.)

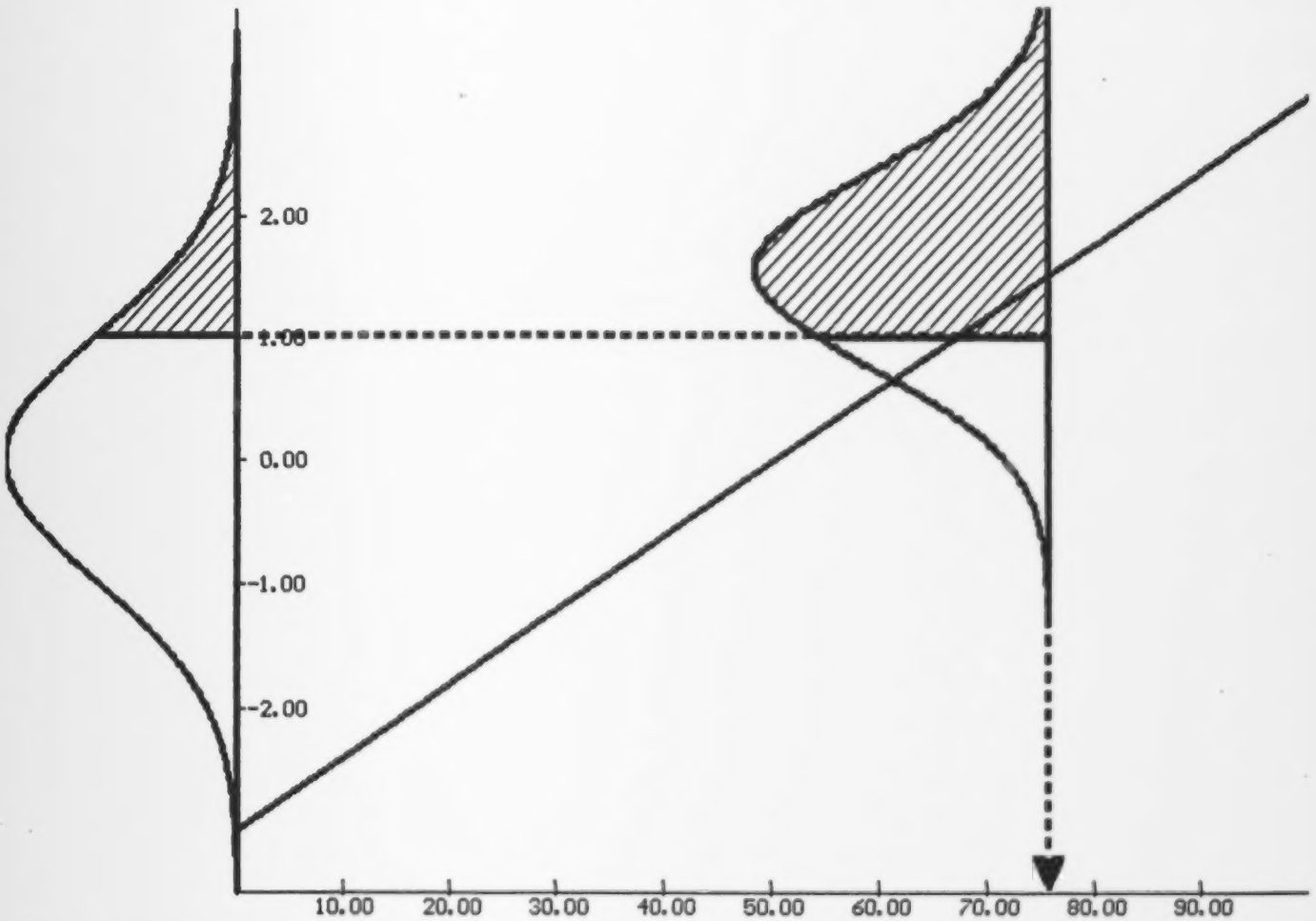
References

Goldman, L. (1972). Tests and counseling: The marriage that failed. Measurement and evaluation in guidance, 4, 213-220.

Magnusson, D. (1967). Test Theory. Palo Alto, CA: Addison Wesley Publishing Company.

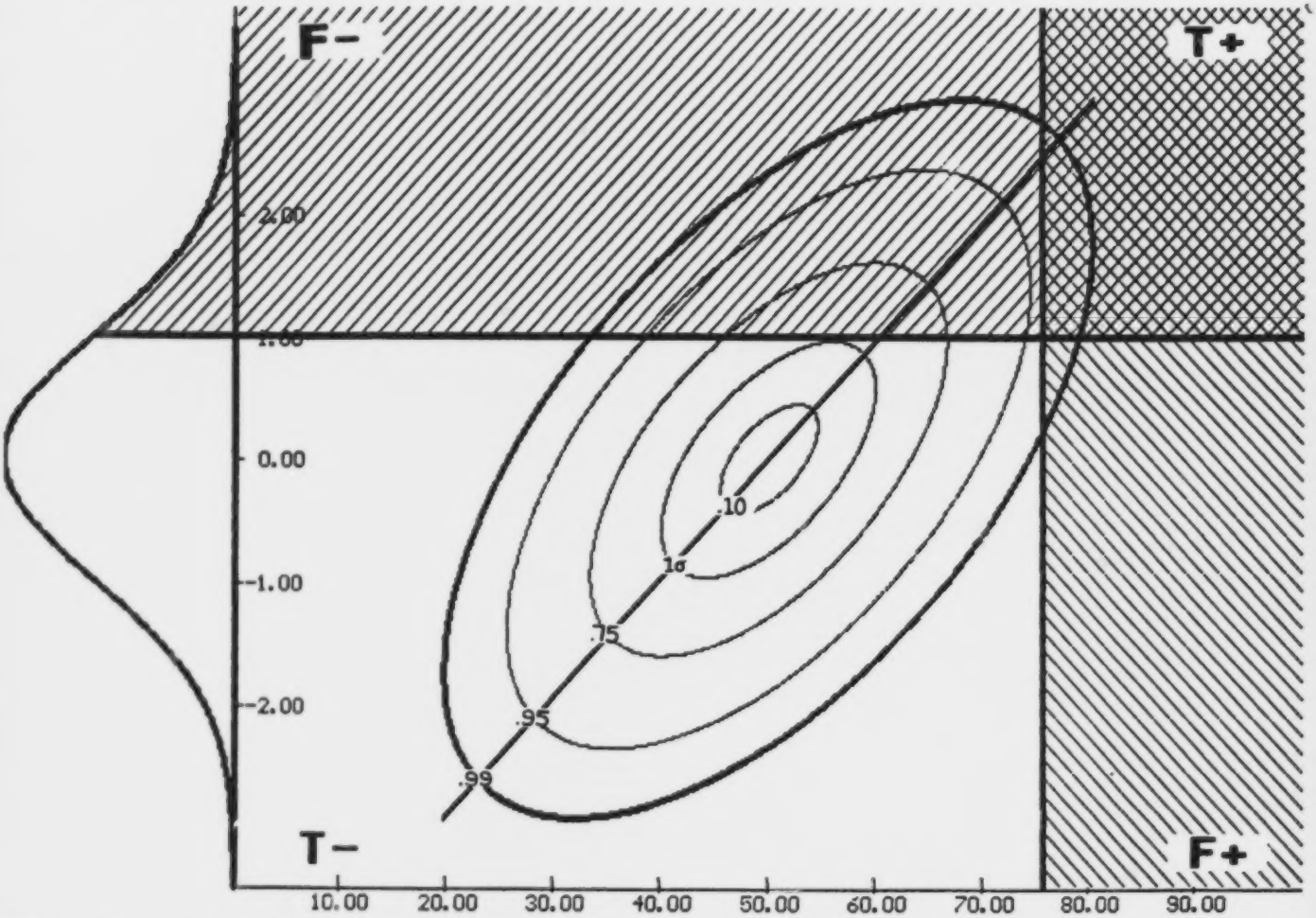
Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. Psychological Bulletin, 52, 194-216.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 22, 565-578.



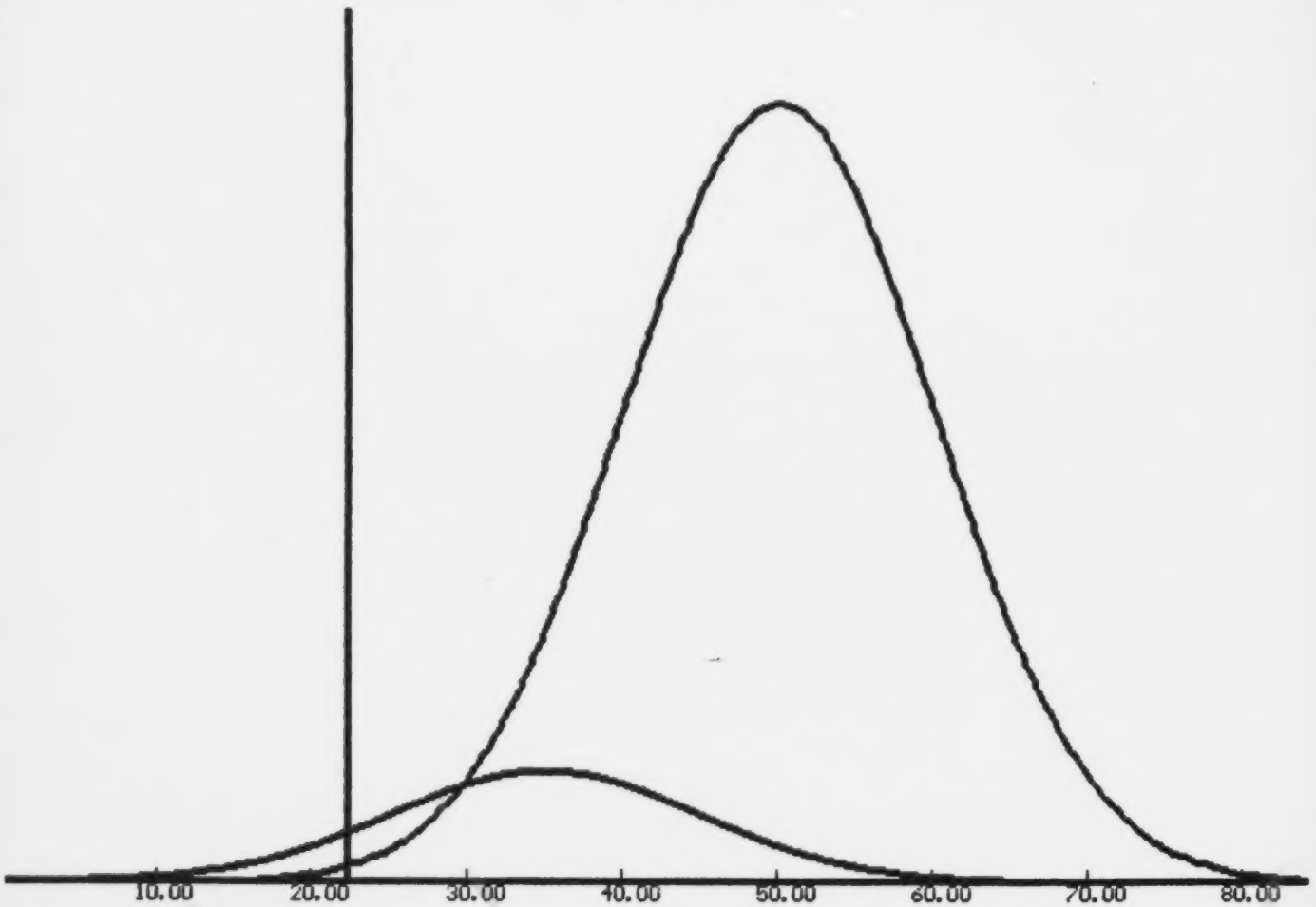
When the correlation is 0.60, if you use a cutoff score of 75.66 on the MMPI D-SCALE to classify clients as DEPRESSED, your maximum probability of error will be 0.25. Do you want to see the long run probabilities associated with this classification rule? (type Y or N, then press Enter):

Figure 1. The cutoff score on the predictor variable when the criterion is normally distributed.



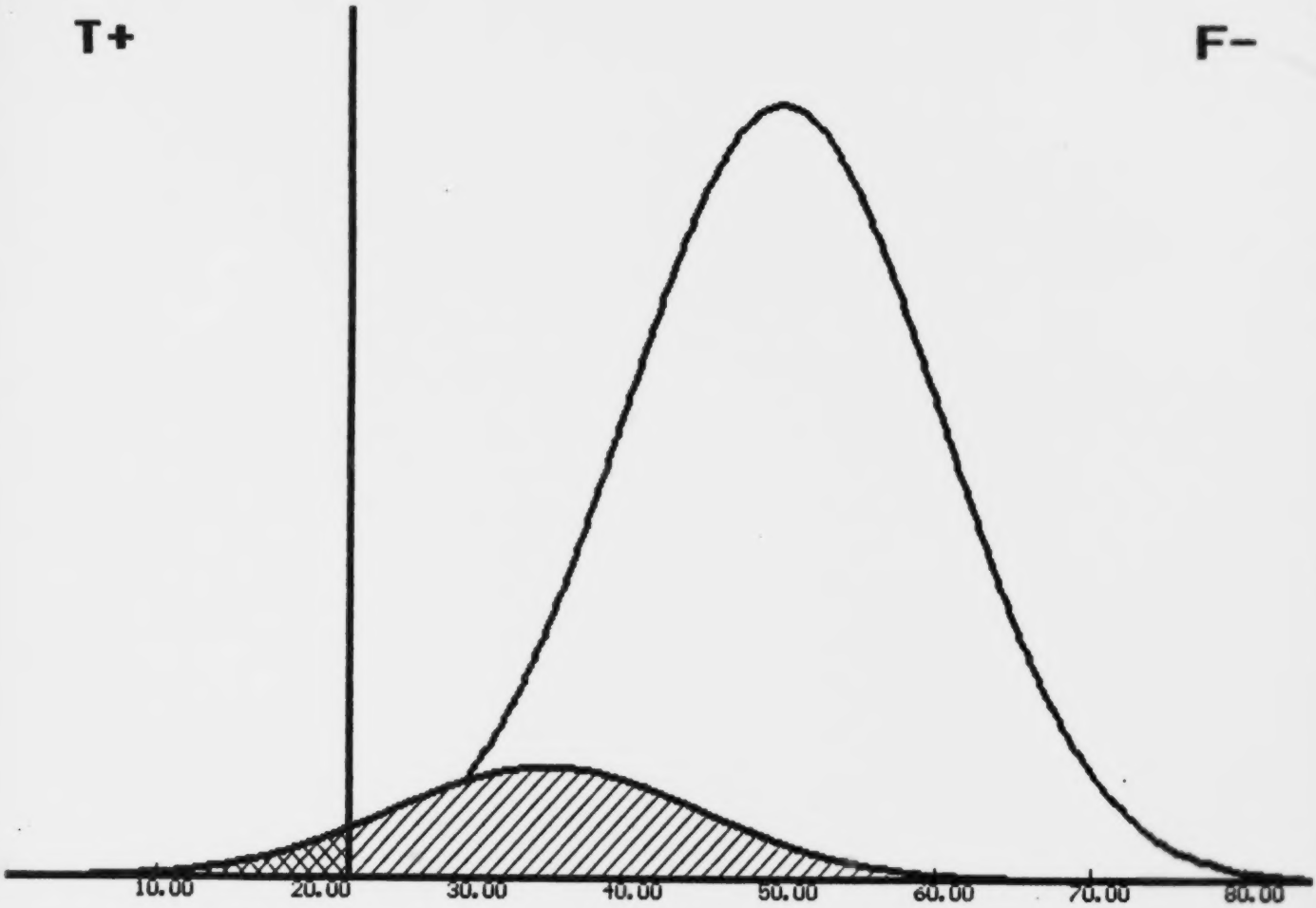
When the correlation is 0.60, in the long run the conditional probability of incorrectly classifying a NOT DEPRESSED client as DEPRESSED is 0.167 (the success rate is 0.833). The proportion of true positives (T+) is 0.004, of false positives (F+) is 0.001, of true negatives (T-) is 0.840, and of false negatives (F-) is 0.154.

Figure 2. The long-run probabilities of true positives, false positives, true negatives, and false negatives when the criterion and predictor have a bivariate normal distribution.



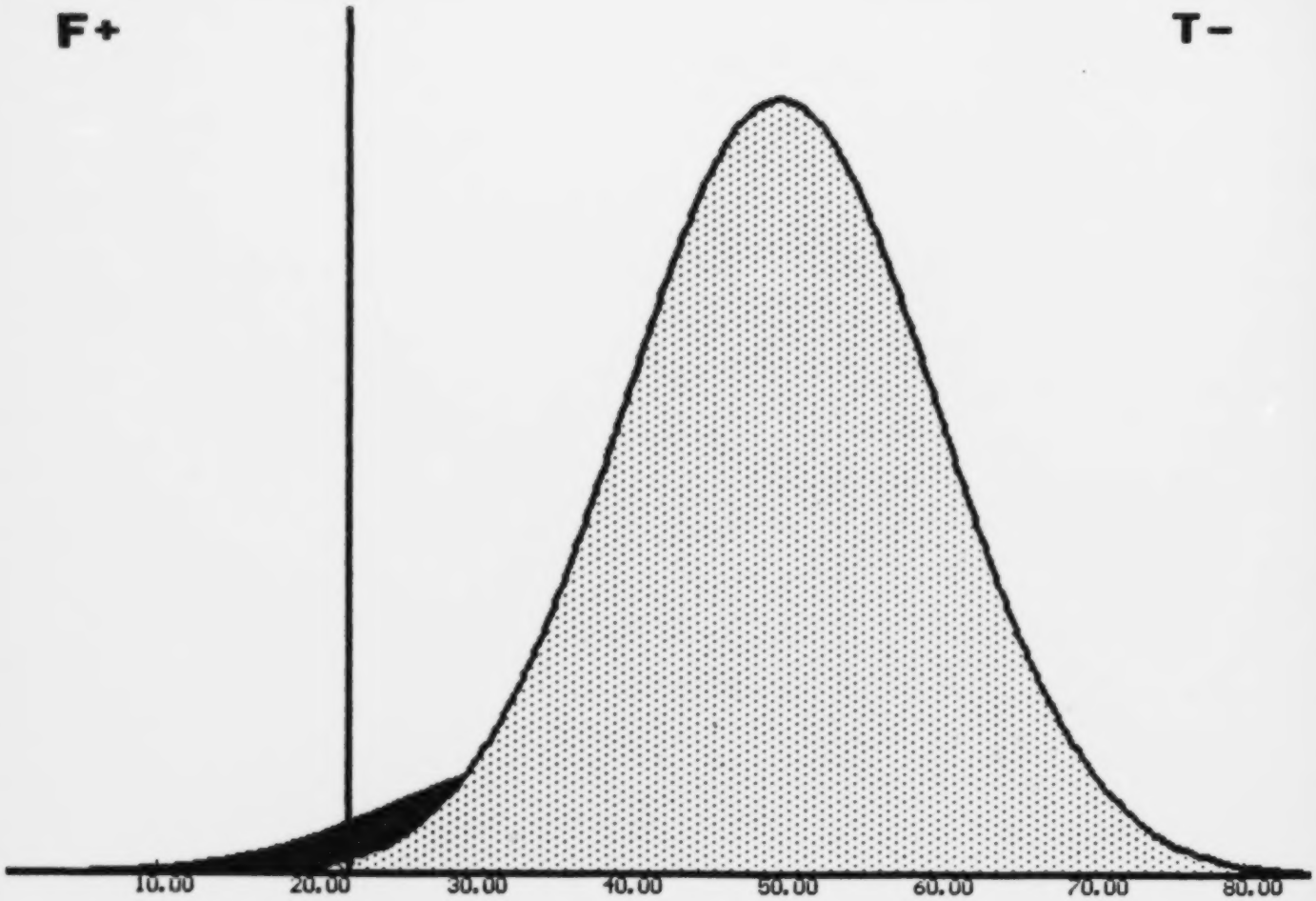
If you use a cutoff score of 22.20 on the CPI SY to classify clients as DELINQUENT, your maximum probability of an error will be 0.25. Do you want to see the long run probabilities associated with this classification rule? (type Y or N, then press Enter):

Figure 3. The cutoff score on a normally distributed predictor variable when the criterion is dichotomous.



The proportion of true positives (T+) is 0.013 and of false negatives (F-) is 0.112.

Figure 4. The long-run probabilities of true positives and false negatives when the predictor variable is normally distributed and the criterion is dichotomous.



The proportion of false positives (F+) is 0.002 and of true negatives (T-) is 0.873. In the long run, the conditional probability of incorrectly classifying a NOT DELINQUENT client as DELINQUENT is 0.159 (the success rate is 0.841).

Figure 5. The long-run probabilities of false positives and true negatives when the predictor variable is normally distributed and the criterion is dichotomous.

APPENDIX

21

Examples of Student Responses to Computer Program

"It put validity information in a very revealing light. Highlighted just how risky it is for a counselor to make predictions. Contributed to a very healthy skepticism."

"I liked the opportunity to change values and receive immediate feedback/ results/outcomes, and thus identify patterns and trends."

"The program is a vivid and dramatic way to highlight the pitfalls of placing too much confidence in test results. Seeing it on the screen is 100 times more effective than talking about it."

"I liked the speed with which it illustrated these concepts."

"It would be difficult to grasp these concepts, I think, by lecture alone."

"I liked being able to experiment with different scenarios."

"No escape without bombing the program. If you went the wrong way in class you had to wait for it to draw contours."

"I would like the program to explain what it is doing, and to show formulas."